PEMODELAN REGRESI HURDLE BINOMIAL NEGATIF UNTUK MENGATASI OVERDISPERSI PADA REGRESI POISSON

SKRIPSI

ANISA HAURA SALSA FATIH YUSUF

H051171513

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

PEMODELAN REGRESI HURDLE BINOMIAL NEGATIF UNTUK MENGATASI OVERDISPERSI PADA REGRESI POISSON

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

ANISA HAURA SALSA FATIH YUSUF

H051171513

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
AGUSTUS 2022

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa Skripsi yang saya buat dengan judul:

Pemodelan Regresi Hurdle Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 22 Agustus 2022

METERAI TEMPEL

ANISA HAURA SALSA FATIH YUSUF

NIM. H051171513

PEMODELAN REGRESI HURDLE BINOMIAL NEGATIF UNTUK MENGATASI OVERDISPERSI PADA REGRESI POISSON

Disetujui Oleh:

Pembimbing Utama,	**Pembimbing Pertama,**
Andi Kresna Jaya, S.Si., M.Si.	**Sitti Sahriman, S.Si., M.Si.**
NIP. 19731228 200003 1001	**NIP. 19881018 201504 2002**

Ketua Departemen Statistika

Dr. Nurtiti Sunusi, S.Si., M.Si.
NIP. 19720117 199703 2022

Pada Tanggal: 22 Agustus 2022

Universitas Hasanuddin

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Anisa Haura Salsa Fatih Yusuf

NIM : H051171513

Program Studi : Statistika

Judul Skripsi : Pemodelan Regresi Hurdle Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

		Tanda Tangan
1. Ketua	: Andi Kresna Jaya, S.Si., M.Si.	(……………………)
2. Sekretaris	: Sitti Sahriman, S.Si., M.Si.	(……………………)
3. Anggota	: Drs. Raupong, M.Si.	(……………………)
4. Anggota	: Anisa, S.Si., M.Si.	(……………………)

Ditetapkan di: Makassar
Tanggal: 22 Agustus 2022

KATA PENGANTAR

Assalamu'alaikum Warohmatullahi Wabarokatuh.

Alhamdulillahi robbil'alamin, Puji syukur kepada **Allah *Subhanahu Wa Ta'ala*** atas segala limpahan rahmat, nikmat, dan hidayah yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul **"Pemodelan Regresi Hurdle Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam.

Salam dan sholawat *InsyaAllah* senantiasa tercurah kepada **Nabi Muhammad *Shallallahu'alaihi Wasallam***, sang kekasih tercinta yang telah memberikan petunjuk cinta dan kebenaran dalam kehidupan.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dari berbagai pihak yang turut membantu baik moril maupun material sehingga akhirnya tugas akhir ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang setinggi-tingginya dan penghargaan yang tak terhingga kepada Ayahanda **Yauri Yusuf** dan Ibunda tercinta **Evy Santy** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran dan dengan limpahan cinta, kasih sayang, dan doa kepada penulis yang tak pernah habis, adik-adik tersayang **Aura Kaula Yusuf, Muhammad Daffa Fatanah Yusuf,** dan **Aqilah Nabilah Zahirah** yang selalu membantu jika ada kendala selama penulisan dan menjadi penyemangat untuk segera menyelesaikan masa studi penulis.

Ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.

3. **Ibu Dr. Nurtiti Sunusi, S.Si., M.Si.**, selaku Ketua Departemen Statistika yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf **Departemen Statistika** yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Bapak Andi Kresna Jaya, S.Si., M.Si.** selaku Pembimbing Utama penulis yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan ditengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis.
5. **Ibu Sitti Sahriman, S.Si., M.Si.** selaku Pembimbing Pertama penulis yang telah meluangkan waktunya ditengah kesibukan untuk memberikan arahan bagi penulis.
6. **Bapak Drs. Raupong, M.Si.** selaku Dosen Penguji sekaligus Penasehat Akademik dan **Ibu Anisa, S.Si., M.Si.** selaku Dosen Penguji, terima kasih telah ikhlas meluangkan waktunya ditengah kesibukan untuk memberikan arahan berupa saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir ini.
7. **Mirna**, **Fakhriyyah**, **Eva**, dan **Kak Fadil** yang telah menjadi teman terbaik sejak awal penulisan skripsi dan senantiasa mendengarkan curhatan, memberikan dorongan, semangat, dan motivasi dalam setiap keadaan sehingga penulis bisa mendapatkan lebih banyak pelajaran hidup.
8. Sahabat terbaik, **Aul**, **Evan**, **Dody**, dan **Etika** yang sampai saat ini masih setia mendengarkan keluh kesah penulis.
9. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka, dan duka selama menjalani pendidikan di Departemen Statistika. Penulis senang mengenal kalian semua, terkhusus **Putri Henri**, **Salsabilah Ramadhani**, **Indry Angelin**, dan **Gabrielle Jesica**.
10. Keluarga besar **DISKRIT 2017**, terima kasih telah memberikan pelajaran yang berharga dan arti kebersamaan selama ini kepada penulis. Pengalaman yang berharga telah penulis dapatkan dari teman-teman selama berproses.

11. **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan di proses perkuliahan dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuddin.
12. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi Allah ***Subhanahu Wa Ta'ala***.

Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga dapat bermanfaat bagi pihak-pihak yang berkepentingan. *Aamiin Yaa Rabbal Alamin.*

Makassar, 22 Agustus 2022

Anisa Haura Salsa Fataih Yusuf

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertandatangan di bawah ini:

Nama	: Anisa Haura Salsa Fatih Yusuf
NIM	: H051171513
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

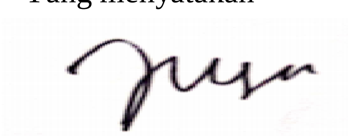
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul:

“**Pemodelan Regresi Hurdle Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson**”

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.
Dibuat di Makassar pada tanggal 22 Agustus 2022

Yang menyatakan



Anisa Haura Salsa Fatih Yusuf

ABSTRAK

Asumsi yang harus dipenuhi dalam regresi Poisson yaitu equidispersi yakni nilai *mean* sama dengan nilai variansi. Namun, dalam beberapa kasus terdapat data yang nilai variansinya lebih besar dari nilai *mean* (overdispersi). Hal ini dapat disebabkan adanya nilai nol berlebih (*excess zeros*) pada data. Model *hurdle* binomial negatif diaplikasikan pada data yang mengalami overdispersi akibat *excess zeros*. Selain itu, paramater dispersi dalam model *hurdle* binomial negatif dapat menjelaskan variansi dalam data. Model *hurdle* terdiri atas model *zero hurdle* dan model *truncated* binomial negatif. Model *zero hurdle* menjelaskan kecenderungan ada atau tidaknya terjadi kasus di suatu lokasi dan model *truncated* binomial negatif menjelaskan seberapa banyak jumlah kasus yang ditemukan di suatu lokasi. Model ini diterapkan pada data jumlah kasus kematian bayi di Kota Makassar Tahun 2017 dengan faktor-faktor yang diduga mempengaruhi diantaranya persentase bayi dengan berat badan lahir rendah, persentase pelayanan kesehatan bayi, persentase bayi yang diberi ASI eksklusif, serta persentase kunjungan neonatus. Pada penelitian ini, regresi *hurdle* binomial negatif diestimasi menggunakan *maximum likelihood estimation* dengan algoritma *Broyden-Fletcher-Goldfarb-Shanno*. Hasil yang diperoleh menunjukkan bahwa model dengan persentase bayi yang diberi ASI eksklusif signifikan mempengaruhi jumlah kasus kematian bayi di Kota Makassar Tahun 2017 baik pada model *zero hurdle* dan model *truncated* binomial negatif. Selain itu, nilai *akaike information criterion minimum* untuk model yang hanya melibatkan variabel persentase bayi yang diberi ASI eksklusif sebesar 112,02.

Kata Kunci: *Hurdle* Binomial Negatif, *Maximum Likelihood Estimation*, Overdispersi, *excess zeros*, Jumlah Kematian Bayi.

ABSTRACT

The Poisson regression assumes equidispersion, which means that the mean value is the same as the variance value. However, in some circumstances, there are data with a variance bigger than the mean value (overdispersion). This can be indicated by the existence of many zeros in the data. The hurdle negative binomial model is applied to data that is overdispersed due to excess zeros. Furthermore, the dispersion parameter in the hurdle negative binomial model can explain the variance in the data. The model consists of a zero hurdle model and a negative binomial truncated model. The negative binomial model describes the enormous number of cases detected in an area, whereas the zero hurdle model explains the probability of cases to exist or not in an area. The Hurdle Negative Binomial regression model was applied to data on the number of infant mortality cases in Makassar City in 2017, with factors assumed to effect infant mortality cases such as the percentage of infants with low birth weight (LBW), the percentage of infant health services, the percentage of infants exclusively breastfed, and the percentage of neonatal visits. In this study, the Hurdle Negative Binomial regression was estimated using the Maximum Likelihood Estimation (MLE) with the Broyden-Fletcher-Goldfarb-Shanno algorithm. The results show that the model with the percentages of exclusively breastfed infants has a significant effect on the number of infant mortality cases in Makassar City in 2017, both on the zero hurdle model and the negative truncated binomial model. Additionally, using a model that just includes the variable percentage of infants that are exclusively breastfed, the minimum akaike information criteria value is 112.02.

Keyword: Hurdle Negative Binomial, Maximum Likelihood Estimation, overdispersion, excess zeros, the number of infant mortality.

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Salah satu indikator derajat kesehatan dan kesejahteraan masyarakat suatu daerah ialah kematian bayi, indikator ini dapat menggambarkan situasional pelayanan kesehatan secara umum di suatu wilayah. Angka Kematian Bayi (AKB) juga dapat menggambarkan kondisi sosial ekonomi masyarakat setempat, karena bayi adalah kelompok usia yang paling rentan terkena dampak dari perubahan lingkungan maupun sosial ekonomi (Abdiana, 2015). Namun, sampai saat ini angka kematian bayi di Sulawesi Selatan masih tergolong tinggi, bahkan Sulawesi Selatan termasuk provinsi dengan jumlah kematian bayi terbanyak di Indonesia sehingga masalah ini perlu diperhatikan. Sepanjang tahun 2017, terdapat 115 kematian bayi di Sulawesi Selatan. Kota Makassar menjadi salah satu dari 5 kabupaten/kota di Sulawesi Selatan dengan tingkat kematian bayi terbanyak (Dinkes, 2018).

Berdasarkan Profil Kesehatan Indonesia dijelaskan bahwa beberapa penyebab kematian bayi dapat disebabkan oleh beberapa variabel antara lain pemberian vitamin A yang tidak terpenuhi, tidak rutin melakukan kunjungan neonatus, pemberian ASI eksklusif pada bayi yang tidak terpenuhi, dan Berat Badan Lahir Rendah (BBLR). Adapun upaya yang dilakukan oleh pemerintah dalam mengantisipasi angka kematian bayi antara lain melalui peningkatan pelayanan kesehatan bayi, kunjungan neonatus, dan sosialisasi mengenai pentingnya imunisasi lengkap bayi (Dinkes, 2018). Oleh karena itu, agar upaya tersebut tepat guna, maka perlu dilakukan pemodelan hubungan jumlah kematian bayi dengan faktor-faktor yang diduga mempengaruhinya.

Hubungan jumlah kematian bayi dengan faktor-faktor penyebabnya dapat diketahui dengan menggunakan salah satu analisis regresi yaitu regresi Poisson. Asumsi penting pada analisis regresi Poisson adalah nilai variansi harus sama dengan nilai rata-rata yang biasa disebut dengan equidispersi (Famoye *et al*., 2004). Namun, asumsi tersebut seringkali tidak terpenuhi yang mana pada kasus tertentu terdapat nilai variansi pada variabel respon lebih kecil dari nilai rata-ratanya (underdispersi) atau nilai variansi data lebih besar dari nilai rata-ratanya (overdispersi). Data yang tidak memenuhi asumsi equidispersi akan mengakibatkan analisis regresi Poisson menghasilkan kesimpulan yang tidak valid. Adapun beberapa hal yang dapat menyebabkan terjadinya overdispersi antara lain banyaknya pengamatan yang bernilai nol (*excess zeros*), adanya sumber keragaman antar-individu yang tidak teramati (*unobserved heteroeneity*), atau adanya pengamatan yang hilang (*data missing*) (Cahyandari, 2014).

Salah satu alternatif yang dapat digunakan untuk mengatasi model regresi Poisson yang mengalami overdispersi yaitu menggunakan model regresi binomial negatif. Model regresi binomial negatif memiliki kegunaan yang sama dengan model regresi Poisson yaitu untuk menganalisis hubungan antara suatu variabel respon data *count* dengan satu atau lebih variabel prediktor, tetapi model regresi binomial negatif lebih fleksibel dibandingkan dengan model Poisson karena asumsi nilai variansi dan nilai rata-ratanya tidak harus sama besar. Model binomial negatif merupakan model campuran antara Poisson dan Gamma yang mana distribusi Gamma digunakan untuk menyesuaikan kehadiran overdispersi dalam model Poisson (Coxe *et al*., 2009).

Pada kasus tertentu, data *count* tidak hanya mengalami overdispersi akan tetapi dapat juga mengalami kelebihan nol (*excess zeros*). Data kelebihan nol (*excess zeros*) adalah kondisi ketika jumlah nilai nol pada variabel respon lebih banyak dari jumlah nilai diskrit lainnya (Pontoh, 2015). Berdasarkan hal tersebut, beberapa solusi yang dapat digunakan untuk mengatasi nilai nol yang berlebih yaitu dengan mengaplikasikan model *zero inflated* atau model *hurdle* (Rahayu *et al*., 2018). *Zero inflated* akan memecah model menjadi 2 bagian, yaitu model pertama merupakan munculnya data *count* yang hanya berisi nilai nol dan model kedua yaitu nilai non-negatif. Model *hurdle* pada dasarnya hampir mirip dengan model *zero inflated* yang melakukan dua pemodelan. Namun, pada pemodelan kedua, model *hurdle* menggunakan *zero truncated* untuk data yang bernilai positif pada data diskrit (Bhakta, 2018). Keunggulan dari model *hurdle* adalah kedua model didalamnya dapat dilakukan penaksiran parameter secara terpisah atau dengan kata lain dimaksimumkan secara terpisah sehingga diharapkan dapat lebih mudah dalam penginterpretasiannya (Cantoni dan Zedini, 2010).

Adapun penelitian-penelitian yang telah dilakukan untuk mengatasi overdispersi dan kelebihan nol diantaranya: Oktari *et al*. (2016) menggunakan regresi binomial negatif untuk mengatasi kasus overdispersi pada data jumlah kematian bayi di Kota Padang, Yulian (2018) memodelkan regresi *zero inflated* binomial negatif untuk mengatasi kasus overdispersi akibat kelebihan nol, dan Hadi (2019) menggunakan regresi *hurdle* Poisson untuk mengatasi kasus overdispersi pada data demam berdarah di Kota Makassar. Namun, menurut Zhang *et al*. (2018) model regresi *hurdle* binomial negatif lebih tepat digunakan untuk mengatasi kasus overdispersi akibat kelebihan nol jika dibandingkan dengan model regresi Poisson, regresi binomial negatif, dan regresi *zero inflated*.

Berdasarkan uraian tersebut, penulis akan melakukan penelitian dengan judul "*Pemodelan Regresi hurdle Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson*" pada data jumlah kematian bayi di Kota Makassar tahun 2017.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mendapatkan estimasi parameter model regresi *hurdle* binomial negatif dengan menggunakan metode *Maximum Likelihood Estimation* (MLE)?
2. Bagaimana memodelkan data jumlah kematian bayi di Kota Makassar tahun 2017 yang mengalami overdispersi dengan menggunakan model regresi *hurdle* binomial negatif?

1.3 Batasan Masalah

Penelitian ini dibatasi pada permasalahan sebagai berikut.

1. Model regresi *hurdle* binomial negatif akan diaplikasikan pada data jumlah kematian bayi di kota Makassar tahun 2017 yang mengalami overdispersi.
2. Metode estimasi parameter model regresi *hurdle* binomial negatif yang digunakan adalah *Maximum Likelihood Estimation* (MLE) dengan algoritma *Broyden-Fletcher-Goldfarb-Shanno* (BFGS).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah maka tujuan dari penelitian ini diuraikan sebagai berikut:

1. Mendapatkan estimasi parameter model regresi *hurdle* binomial negatif dengan menggunakan metode *Maximum Likelihood Estimation* (MLE).
2. Mendapatkan dugaan model *hurdle* binomial negatif pada data jumlah kematian bayi di Kota Makassar tahun 2017 yang mengalami overdispersi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan wawasan keilmuan yang lebih khusus kepada penulis tentang pemodelan regresi *hurdle* binomial negatif untuk data diskrit yang mengalami masalah overdispersi.
2. Memberikan informasi kepada pemerintah untuk menetapkan kebijakan dalam rangka mengurangi Angka Kematian Bayi (AKB) di Kota Makassar.

BAB II
TINJAUAN PUSTAKA

2.1 *Generalized Linier Model*

Mc Cullagh dan Nelder (1989) mendefinisikan bahwa *Generealized Linier Models* (GLM) merupakan perluasan dari model linier klasik. Sifat penting dari model ini adalah asumsi normalitas dan kehomogenan variansi tidak diperlukan sehingga hubungan antara variabel respon dan prediktor dengan distribusi keluarga eksponensial dapat dimodelkan dengan model ini. Agresti (2002) menyatakan ada tiga komponen dalam GLM, yaitu:

1. Komponen Acak

   Variabel respon $Y = (y_1, y_2, \dots, y_n)$ dalam suatu n observasi diasumsikan saling bebas dan memiliki distribusi yang termasuk dalam keluarga eksponensial, dengan fungsi kepadatan peluang sebagai berikut :

   $$f(y_i;\theta_i,\phi) = \exp\left\{\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i;\phi)\right]\right\} \tag{2.1}$$

   Parameter θ_i disebut dengan parameter natural dengan $i = 1,2,\dots,n$. Parameter ϕ disebut parameter dispersi. Jika ϕ konstan yang diketahui, maka Persamaan (2.1) dinyatakan dalam bentuk :

   $$f(y_i;\theta_i) = \exp\{y_i\theta_i - b(\theta_i) + c(y_i)\} \tag{2.2}$$

2. Komponen Sistematis

   Kombinasi linier variabel prediktor $\boldsymbol{X}$ dengan koefisien parameter $\boldsymbol{\beta}$ yang menghasilkan penduga linier $\boldsymbol{\eta}$, yaitu $\boldsymbol{\eta} = \boldsymbol{X\beta}$**.** Masing-masing dari elemen $\boldsymbol{\eta}$ dapat dinyatakan dengan persamaan sebagai berikut:

   $$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \tag{2.3}$$

   $$\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}, \text{ dengan } i = 1,2,\dots,n \tag{2.4}$$

   atau dalam bentuk matriks

$$\begin{bmatrix} \eta_i \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

dengan n merupakan banyaknya amatan dan p merupakan banyaknya variabel prediktor.

3. Fungsi Penghubung

   Fungsi penghubung $g(\cdot)$ adalah suatu fungsi yang menghubungkan komponen acak dengan komponen sistematis. Diketahui $\mu_i = E(y_i)$, maka model yang menghubungkan μ_i dengan prediktor linier η_i dapat dinyatakan dengan persamaan sebagai berikut:

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}, \text{ dengan } i = 1,2, \dots, n \qquad (2.5)$$

   dengan fungsi g menunjukkan fungsi penghubung.

2.2 Regresi Poisson

Menurut Cahyandari (2014), regresi Poisson merupakan salah satu dari model regresi yang berasal dari distribusi Poisson yang biasanya digunakan untuk menganalisis data dengan variabel respon berupa data diskrit yang nilainya berupa bilangan bulat non-negatif. Misalkan y_i, $i = 1,2, \dots, n$ merupakan jumlah kejadian yang muncul dalam selang waktu dengan rata-rata μ_i. Jika Y adalah variabel acak Poisson dengan parameter $\mu > 0$, maka fungsi kepadatan peluangnya adalah (Agresti, 2002):

$$f(y_i; \mu) = \frac{e^{-\mu} \mu^{y_i}}{y_i!} \; y_i = 0,1,2, \dots \qquad (2.6)$$

Diketahui dalam GLM, kombinasi linier variabel prediktor X dengan koefisien parameter β yang menghasilkan penduga linier η, yaitu :

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \qquad (2.7)$$

Pada distribusi Poisson diketahui bahwa $\mu_i = E(y_i)$. Model yang menghubungkan μ_i dengan prediktor linier η_i, dinyatakan dengan $g(\mu_i) = \eta_i$.

Jika *ln link* yang digunakan, maka berdasarkan konsep GLM untuk distribusi Poisson, hubungan μ_i dengan prediktor linier η_i dinyatakan dengan $\ln(\mu_i) = \eta_i$. Maka dengan menggunakan fungsi penghubung log natural tersebut diperoleh model regresi Poisson dalam bentuk:

$$\ln \mu_i = \eta_i$$

$$\ln \mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right) \tag{2.8}$$

Pada regresi Poisson terdapat asumsi yang harus dipenuhi, yaitu nilai varians dan nilai *mean* dari variabel respon harus memiliki nilai yang sama atau asumsi equidispersi yang ditunjukkan pada persamaan berikut (Cameron & Trivedi, 1998):

$$E(Y_i) = V(Y_i) = \mu_i \tag{2.9}$$

Model regresi Poisson dengan penghubung log natural dapat ditulis sebagaimana Persamaan (2.10) (Cahyandari, 2014)

$$y_i = \mu_i + \varepsilon_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right) + \varepsilon_i \tag{2.10}$$

dengan

y_i = jumlah kejadian ke-i

μ_i = *mean* jumlah kejadian

ε_i = galat *error* , $i = 1,2, \dots, n$.

2.3 Overdispersi

Salah satu asumsi yang harus dipenuhi dalam regresi Poisson adalah asumsi equidispersi, yaitu keadaan dengan nilai *mean* dan nilai variansi dari variabel responnya bernilai sama. Namun, asumsi pada regresi Poisson seringkali mengalami penyimpangan diantaranya overdispersi yaitu nilai variansi pada data lebih besar dari nilai *mean*-nya, dan underdispersi yaitu keadaan dengan nilai *mean* pada data lebih besar dari nilai variansinya. Pada kasus regresi Poisson yang mengalami overdispersi jika tetap dilanjutkan maka akan diperoleh estimator yang konsisten akan tetapi tidak efisien, standar *error* menjadi *underestimate*, maka kesimpulan yang dihasilkan tidak valid. Adapun salah satu penyebab terjadinya overdispersi yaitu *excees zeros* pada variabel respon. Pemeriksaan overdispersi dapat dilakukan dengan menggunakan nilai *deviance*. Overdispersi dideteksi menggunakan nilai *deviance* dibagi dengan derajat bebas yang mempunyai nilai dispersi lebih besar dari 1 (Hilbe, 2011). Menurut Agresti (2002), nilai *deviance* dapat dinyatakan sebagaimana Persamaan (2.11).

$$\text{dispersi} = \frac{dev}{db}$$

$$dev = 2\sum_{i=1}^{n}\left\{y_i \log\frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i)\right\} \tag{2.11}$$

dengan :

dev : nilai *deviance*

db: derajat bebas, db $= n - (p + 1)$

y_i: nilai variabel respon ke-*i* dengan $i = 1,2, \ldots, n$

$\hat{y}_i$: estimasi regresi Poisson

2.4 Regresi Binomial Negatif

Pada kasus overdispersi diperlukan suatu distribusi yang lebih fleksibel dibandingkan Poisson. Distribusi yang sering digunakan untuk kasus overdispersi adalah binomial negatif. Model binomial negatif tidak menekankan pada sifat equidispersi. Model binomial negatif merupakan salah satu solusi untuk mengatasi masalah overdispersi yang didasarkan pada model campuran Poisson-Gamma (Hardin & Hilbe, 2007). Fungsi kepadatan peluang dari model binomial negatif adalah sebagai berikut:

$$f(y;\mu,\phi) = \frac{\Gamma\left(y + \frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right) y!}\left(\frac{1}{1+\phi\mu}\right)^{1/\phi}\left(\frac{\phi\mu}{1+\phi\mu}\right)^{y} \tag{2.12}$$

dengan y merupakan nilai data diskrit, μ adalah nilai $E(y)$ dan ϕ merupakan paramer dispersi. Kondisi overdispersi ditunjukkan dengan nilai $\phi > 1$.

Misalkan ingin mengetahui hubungan antara suatu variabel respon Y dengan p buah variabel prediktor $x_1, x_2, \dots, x_p$ dengan y_i merupakan data diskrit dan menyatakan banyaknya kejadian yang diamati pada suatu populasi tertentu. Variabel Y diasumsikan berdistribusi binomial negatif. Diberikan sampel acak berukuran n yaitu $\left\{\left(y_i, \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)\right); i = 1,2, \dots, n\right\}$ dengan y_i adalah pengamatan ke-i dari variabel respon Y dan $x_{i1}, x_{i2}, \dots, x_{ip}$ adalah pengamatan ke-i dari variabel prediktor $x_1, x_2, \dots, x_p$. Berikut merupakan model regresi secara umum.

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \ ; i = 1,2, \dots, n \tag{2.13}$$

dengan $\beta_0, \beta_1, \dots, \beta_p$ menyatakan parameter yang tidak diketahui dan ε_i menyatakan *error* untuk pengamatan ke-i dan diasumsikan nilai $E(\varepsilon_i) = 0$.

Misal diasumsikan nilai ekspektasi untuk y_i adalah $E(y_i) = \mu_i$ dan sebelumnya telah diasumsikan bahwa nilai $E(\varepsilon_i) = 0$, maka akan diperoleh:

$$E(y_i) = \mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \tag{2.14}$$

Nilai rataan dari variabel acak adalah diskrit dan bernilai positif. Maka nilai $E(y_i)$ tidak mungkin negatif. Untuk mengatasi keadaan tersebut maka diperlukan suatu fungsi penghubung $g(\mu_i)$ yang menghubungkan antara μ_i dengan prediktor linier yaitu (Fatmasari, 2014):

$$g(\mu_i) = \ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right) \tag{2.15}$$

2.5 Regresi Hurdle

Salah satu pendekatan yang dapat digunakan untuk mengatasi overdispersi adalah model *hurdle*. Overdipersi terjadi ketika nilai variansi lebih besar dari nilai *mean*. Salah satu penyebab terjadi overdispersi adalah banyaknya nilai nol pada variabel respon (*excess zero*). Model *hurdle* mampu mengatasi kasus *excess zeros* dengan membagi dua model ke dalam dua bagian yaitu:

1. Model untuk data biner yang bernilai nol (*zero counts*) atau nilai positif (*positive counts*), yang mana model tersebut ditaksir dengan model logistik.
2. Model untuk data yang bernilai positif (*positive counts*), yang mana model tersebut ditaksir dengan *truncated model*.

Misalkan $k_1(0)$ adalah fungsi probabilitas ketika variabel respon (Y) bernilai sama dengan 0 dan $k_2(y)$ dengan $y = 1,2,3, \dots$ adalah fungsi probabilitas ketika variabel respon (Y) bernilai non-negatif sehingga fungsi probabilitas dari model *hurdle* adalah sebagai berikut (Saffari *et al*., 2012):

$$f(y) = \begin{cases} k_1(0), & y_i = 0 \\ \big(1 - k_1(0)\big)k_2(y), & y_i > 0 \end{cases} \tag{2.16}$$

Mullahy (1986) menjelaskan tentang dua model *hurdle* yang terbentuk pada Persamaan (2.16) yang dimisalkan sebagai $f_1(y)$ dan $f_2(y)$, yang mana $k_1(0) = f_1(0)$ dan pada $k_2(y)$ akan dilakukan normalisasi karena $f_2(y)$ berlaku untuk nilai non-negatif dengan $y = 0,1,2, \dots$ sedangkan $k_2(y)$ berlaku untuk nilai positif dengan $y = 1,2,3, \dots$ yang dapat didefinisikan sebagai berikut:

$$f_2(0) + f_2(1) + f_2(2) + \cdots = 1$$

$$f_2(0) - f_2(0) + f_2(1) + f_2(2) + \cdots = 1 - f_2(0)$$

$$f_2(1) + f_2(2) + \cdots = 1 - f_2(0)$$

$$\frac{f_2(1)}{1-f_2(0)} + \frac{f_2(2)}{1-f_2(0)} + \cdots = 1$$

Maka untuk $y = 1,2, \dots$ berlaku $k_2(y) = \frac{f_2(y)}{1-f_2(0)}$ sehingga distribusi probabilitas dari model regresi *hurdle* adalah sebagai berikut.

$$f(y) = \begin{cases} f_1(0), & y_i = 0 \\ \dfrac{1 - f_1(0)}{1 - f_2(0)} f_2(y), & y_i > 0 \end{cases} \tag{2.17}$$

2.6 Regresi Hurdle Binomial Negatif

Regresi *Hurdle* Binomial Negatif (HBN) digunakan untuk variabel respon berupa data diskrit, memiliki nilai nol dengan proporsi lebih besar dari nilai lainnya (*excess zero*), dan mengalami overdispersi (Desjardins, 2013). Misalkan $y_i(i = 1.2, \dots, n)$ merupakan variabel respon diskrit yang diperoleh dari data cacah sebanyak *n* dari pengukuran observasi ke-*i*. Berikut adalah metode menentukan distribusi probabilitas dari model regresi HBN yang dibentuk berdasarkan Persamaan (2.17):

$$f_1(0) = \pi \tag{2.18}$$

$$f_2(y_i) = \frac{\Gamma\left(y_i + \frac{1}{\phi}\right)}{\Gamma(y_i + 1)\Gamma\left(\frac{1}{\phi}\right)}\left(\frac{\phi\mu}{1 + \phi\mu}\right)^{y_i}(1 + \phi\mu)^{-\phi^{-1}} \tag{2.19}$$

dan

$$\begin{aligned} f_2(0) &= \frac{\Gamma\left(0 + \frac{1}{\phi}\right)}{\Gamma(0 + 1)\Gamma\left(\frac{1}{\phi}\right)}\left(\frac{\phi\mu}{1 + \phi\mu}\right)^0(1 + \phi\mu)^{-\phi^{-1}} \\ &= \frac{\Gamma\left(\frac{1}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)}(1 + \phi\mu)^{-\phi^{-1}} \\ &= (1 + \phi\mu)^{-\phi^{-1}} \end{aligned} \tag{2.20}$$

dari Persamaan (2.18),(2.19), dan (2.20) diperoleh:

$$\begin{aligned} \frac{1-f_1(0)}{1-f_2(0)}f_2(y_i) &= \frac{1-\pi}{1-(1+\phi\mu)^{-\phi^{-1}}}\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\phi}\right)}\left(\frac{\phi\mu}{1+\phi\mu}\right)^{y_i}(1 + \phi\mu)^{-\phi^{-1}} \\ &= (1-\pi)\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma(y_i+1)\Gamma\frac{1}{\phi}}\left(\frac{\phi\mu}{1+\phi\mu}\right)^{y_i}\frac{(1+\phi\mu)^{-\phi^{-1}}}{1-(1+\phi\mu)^{-\phi^{-1}}} \end{aligned} \tag{2.21}$$

sehingga diperoleh fungsi kepadatan peluang untuk model regresi *hurdle* binomial negatif adalah sebagai berikut (Saffari *et al*., 2012):

$$P(Y_i = y_i) = \begin{cases} \pi, & y_i = 0 \\ (1-\pi)\frac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma(y_i+1)\Gamma\frac{1}{\phi}}\left(\frac{\phi\mu}{1+\phi\mu}\right)^{y_i}\frac{(1+\phi\mu)^{-\phi^{-1}}}{1-(1+\phi\mu)^{-\phi^{-1}}}, & y_i > 0 \end{cases} \tag{2.22}$$

Nilai variabel respon pada pengamatan muncul dalam dua keadaan yang terpisah. Keadaan pertama disebut *zero hurdle* terjadi dengan peluang π, sementara keadaan kedua disebut *truncated* binomial negatif terjadi dengan peluang $1 - \pi$, dengan $0 < \pi < 1$, μ adalah *mean* dari distribusi binomial negatif dengan $\mu > 0$ dan ϕ adalah parameter dispersi yang tidak bergantung pada variabel prediktor dengan $\phi > 0$. Diketahui bahwa π dan μ bergantung pada vektor dari variabel prediktor yang dapat didefinisikan sebagai berikut:

a. Model *zero hurdle* (model logistik)

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j=1}^{p} z_{ij}\delta_j$$

$$\frac{\pi_i}{1-\pi_i} = \exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)$$

$$\pi_i = (1 - \pi_i)\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)$$

$$\pi_i = \exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right) - (\pi_i)\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)$$

$$\pi_i\left(1 + \exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)\right) = \exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)$$

maka diperoleh:

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)}{1 + \exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)} \tag{2.23}$$

dengan $0 < \pi_i < 1$, $i = 1,2,\dots,n$, dan $j = 1,2,\dots,p$.

Penduga model untuk *zero hurdle* dengan fungsi penghubung logit dinyatakan sebagai berikut:

$$\text{logit}(\pi_i) = \hat{\delta}_0 + \sum_{j=1}^{p} z_{ij}\,\hat{\delta}_j \tag{2.24}$$

Apabila disajikan dalam bentuk matriks, Persamaan (2.24) dapat ditulis

$$\begin{bmatrix} logit\,\pi_1 \\ logit\,\pi_2 \\ \vdots \\ logit\,\pi_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & \dots & z_{1p} \\ 1 & z_{21} & \dots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \dots & z_{np} \end{bmatrix} \begin{bmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \\ \vdots \\ \hat{\delta}_p \end{bmatrix}$$

b. Model *truncated* binomial negatif

Selanjutnya, nilai μ didapatkan dari model log natural sebagai berikut:

$$ln(\mu_i) = \sum_{j=1}^{p} x_{ij}\beta_j$$

$$\mu_i = \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) \tag{2.25}$$

dengan $i = 1,2,\ldots,n$, dan $j = 1,2,\ldots,p$.

Penduga model untuk *truncated negative binomial* dengan fungsi penghubung log natural dinyatakan sebagai berikut:

$$\ln(\mu_i) = \hat{\beta}_0 + \sum_{j=1}^{p} x_{ij}\,\hat{\beta}_j \tag{2.26}$$

Apabila disajikan dalam bentuk matriks, Persamaan (2.26) dapat ditulis:

$$\begin{bmatrix} ln\ \mu_1 \\ ln\ \mu_2 \\ \vdots \\ ln\ \mu_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

Kemudian, berdasarkan fungsi probabilitas y_i dari Persamaan (2.22) nilai π_i dan μ_i disubtitusikan dari Persamaan (2.23) dan (2.25), maka diperoleh fungsi probabilitas model regresi HBN sebagaimana Persamaan (2.27)

$$P(Y_i = y_i) = \begin{cases} \dfrac{\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)}{1+\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)} & y_i = 0 \\ \dfrac{1}{1+\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)} \dfrac{\Gamma\left(y_i+\frac{1}{\phi}\right)}{\Gamma(y_i+1)\Gamma\frac{1}{\phi}} \left(\dfrac{\phi\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)}{1+\phi\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)}\right)^{y_i} \dfrac{\left(1+\phi\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)\right)^{-\phi^{-1}}}{1-\left(1+\phi\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)\right)^{-\phi^{-1}}} & ,y_i > 0 \end{cases} \tag{2.27}$$

2.7 Pengujian Parameter Regresi Hurdle Binomial Negatif

Setelah mendapatkan model, untuk memeriksa peranan variabel-variabel prediktor dalam model, perlu dilakukan pengujian terhadap parameter model (δ_j dan β_j, dengan $j = 1,2,\ldots,p$). Pengujian terhadap parameter model dilakukan baik secara simultan maupun secara parsial. Pengujian parameter secara simultan dengan menggunakan uji *Likelihood Ratio* dan pengujian parameter secara parsial dilakukan menggunakan statistik uji *Wald*.

2.7.1 Pengujian Simultan

Statistik uji G adalah uji *likelihood ratio* yang digunakan untuk mengetahui variabel prediktor yang secara bersama-sama mempengaruhi variabel respon secara signifikan. Hipotesis yang digunakan adalah sebagai berikut (Kurniawan, 2017).:

Model *zero hurdle*

a. Hipotesis

H_0 : $\delta_1 = \delta_2 = \cdots = \delta_p = 0$ (variabel prediktor secara bersama-sama tidak memiliki pengaruh terhadap variabel respon)

H_1 : Ada $\delta_j \neq 0, j = 1,2, \ldots, p$ (variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon)

b. Statistik Uji

$$G_{hit} = -2 \ln \left[\frac{L_0(\hat{\delta})}{L_1(\hat{\delta})} \right] = -2\left[\ln L_0(\hat{\delta}) - \ln L_1(\hat{\delta}) \right] \tag{2.28}$$

dengan L_0 adalah nilai fungsi *likelihood* untuk model yang tidak mengandung variabel prediktor dan L_1 adalah nilai fungsi *likelihood* untuk model yang mengandung semua variabel prediktor.

c. Kriteria Pengujian

H_0 ditolak jika $G_{hit} > \chi^2_{(\alpha,p)}$ dengan p adalah banyaknya variabel prediktor.

d. Kesimpulan

Penolakan terhadap H_0 menunjukkan bahwa variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon.

Model *truncated* binomial negatif

a. Hipotesis

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ (variabel prediktor secara bersama-sama tidak memiliki pengaruh terhadap variabel respon)

H_1 : Ada $\beta_j \neq 0, j = 1,2, \ldots, p$ (variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon)

b. Statistik Uji

$$G_{hit} = -2 \ln \left[\frac{L_0(\hat{\beta})}{L_1(\hat{\beta})}\right] = -2\left[\ln L_0(\hat{\beta}) - \ln L_1(\hat{\beta})\right] \quad (2.29)$$

dengan L_0 adalah nilai fungsi *likelihood* untuk model yang tidak mengandung variabel prediktor dan L_1 adalah nilai fungsi *likelihood* untuk model yang mengandung semua variabel prediktor.

c. Kriteria Pengujian

H_0 ditolak jika $G_{hit} > \chi^2_{(\alpha,p)}$ dengan *p* adalah banyaknya variabel prediktor.

d. Kesimpulan

Penolakan terhadap H_0 menunjukkan bahwa variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon.

2.7.2 Pengujian Parsial Satu-Satu

Pengujian parameter regresi secara parsial dilakukan menggunakan statistik uji *Wald*, statistik uji ini digunakan untuk menguji signifikansi parameter regresi secara parsial pada masing-masing variabel prediktor. Pengujian parameter secara parsial untuk masing-masing model *zero hurdle* dan *truncated* binomial negatif sebagai berikut (Kurniawan, 2017):

Model *zero hurdle*

a. Hipotesis

$H_0: \delta_j = 0$ (variabel prediktor ke-j tidak memiliki pengaruh terhadap variabel respon)

$H_1:\ \delta_j \neq 0$ dengan $j = 1,2,\dots,p$ (variabel prediktor ke-j memiliki pengaruh terhadap variabel respon)

b. Statistik Uji

$$W(\hat{\delta}_j) = \left(\frac{\hat{\delta}_j}{SE(\hat{\delta}_j)}\right)^2 \quad (2.30)$$

dengan $\hat{\delta}_j$ adalah estimator dari δ_j dan $SE(\hat{\delta}_j)$ yaitu *standard error* dari estimator $\hat{\delta}_j$ pada model *zero hurdle*.

c. Kriteria Pengujian

H_0 ditolak jika $W_j > \chi^2_{(\alpha,1)}$

d. Kesimpulan

Penolakan terhadap H_0 menunjukkan bahwa variabel prediktor ke-j memiliki pengaruh terhadap variabel respon pada tingkat signifikan α.

Model *truncated* binomial negatif

a. Hipotesis

H_0: $\beta_j = 0$ (variabel prediktor ke-j tidak memiliki pengaruh terhadap variabel respon)

H_1: $\beta_j \neq 0$ dengan $j = 1,2,\dots,p$ (variabel prediktor ke-j memiliki pengaruh terhadap variabel respon)

b. Statistik Uji

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \tag{2.31}$$

dengan $\hat{\beta}_j$ adalah estimator dari β_j dan $SE(\hat{\beta}_j)$ yaitu *standard error* dari estimator $\hat{\beta}_j$ pada model *truncated* binomial negatif.

c. Kriteria Pengujian

H_0 ditolak jika $W_j > \chi^2_{(\alpha,1)}$

d. Kesimpulan

Penolakan terhadap H_0 menunjukkan bahwa variabel prediktor ke-j memiliki pengaruh terhadap variabel respon pada tingkat signifikan α.

2.8 Algoritma *Broyden-Fletcher-Goldfarb-Shanno*

Estimasi parameter model HBN dilakukan dengan metode *maximum likelihood estimation* (MLE). Metode MLE digunakan untuk mendapatkan penyelesaian maksimum dari fungsi *likelihood*, namun hanya dapat diperoleh sistem persamaan non-linier. Sistem persamaan non-linier dapat diselesaikan menggunakan pendekatan numerik, yaitu metode *Newton Raphson*. Metode *Newton Raphson* digunakan untuk menyelesaikan sistem persamaan non-linier hingga diperoleh nilai parameter yang maksimum. Namun pada sistem persamaan non-linier tidak dapat dijamin konvergensinya oleh karena itu dapat digunakan metode *Quasi Newton* yaitu algoritma *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) (Saputro dan Widyaningsih, 2016).

Algoritma BFGS merupakan modifikasi dari metode *Newton Raphson* dengan mengganti matriks Hessian pada metode *Newton Raphson* yang menggunakan nilai dari derivatif kedua fungsi log *likelihood* menjadi matriks Hessian yang dihitung menggunakan informasi yang didapatkan pada saat iterasi berlangsung. Untuk memperoleh estimasi parameter menggunakan metode BFGS digunakan persamaan sebagai berikut (Wei *et al*., 2006):

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \lambda^{(k)}\left(\boldsymbol{H}^{(k)}\right)^{-1}\boldsymbol{g}^{(k)} \tag{2.32}$$

dengan:

$\lambda^{(k)} = \min \ln L\left(\boldsymbol{x}^{(k)} - \lambda\left(\boldsymbol{H}^{(\boldsymbol{k})}\right)^{-1}\boldsymbol{g}^{(k)}\right)$

$\boldsymbol{g}$ merupakan derivatif pertama dari fungsi log *likelihood* yaitu $\nabla f\left(\boldsymbol{x}^{(k)}\right)$.

$\boldsymbol{H}$ merupakan matriks Hessian yang diperoleh dengan persamaan:

$$\boldsymbol{H}^{(k+1)} = \left(\boldsymbol{I} - \frac{\boldsymbol{s}^{(k)}\left(\boldsymbol{y}^{(k)}\right)^T}{\left(\boldsymbol{y}^{(k)}\right)^T\boldsymbol{s}^{(k)}}\right)\boldsymbol{H}^{(k)}\left(\boldsymbol{I} - \frac{\boldsymbol{y}^{(k)}\left(\boldsymbol{s}^{(k)}\right)^T}{\left(\boldsymbol{y}^{(k)}\right)^T\boldsymbol{s}^{(k)}}\right) + \frac{\boldsymbol{s}^{(k)}\left(\boldsymbol{s}^{(k)}\right)^T}{\left(\boldsymbol{y}^{(k)}\right)^T\boldsymbol{s}^{(k)}} \tag{2.33}$$

dengan:

$\boldsymbol{s}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$

$\boldsymbol{y}^{(k)} = \boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}$

$\boldsymbol{H}^{(0)} =$ matriks identitas $n \times n$

Persamaan akan terus berulang hingga diperoleh estimasi parameter yang konvergen yaitu pada saat $\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right\| \leq \varepsilon$, yang mana ε merupakan bilangan yang sangat kecil mendekati nol yaitu 10^{-3}.

2.9 Pemilihan Model Terbaik

Menurut Sharma dan Landage (2013), salah satu kriteria untuk menentukan model terbaik adalah *Akaike Information Criteration* (AIC). Dengan kriteria AIC, model terbaik dipilih dengan mempertimbangkan jumlah parameter dalam dalam model. Kriteria AIC mampu menunjukkan seberapa tepat model tersebut dengan data yang dimiliki secara mutlak. Kriteria AIC didefinisikan sebagai berikut.

$$AIC = 2p - 2\ln L(\hat{\theta}) \tag{2.34}$$

dengan $L(\hat{\theta})$ adalah nilai fungsi *likelihood* dan *p* adalah banyaknya parameter.

2.10 Kematian Bayi

Kematian bayi adalah kematian yang terjadi saat setelah bayi lahir sampai bayi belum berusia tepat satu tahun. Angka Kematian Bayi (AKB) dihitung per 1.000 kelahiran hidup pada tahun tertentu. Adapun faktor yang menjadi penyebab kematian bayi, yaitu kematian bayi endogen atau kematian neonatal, dan kematian eksogen atau kematian *post* neonatal (Dinkes, 2018).

Kematian neonatal adalah banyaknya kematian bayi yang terjadi pada bulan pertama setelah dilahirkan dan umumnya disebabkan oleh faktor yang dibawa anak sejak lahir yang diperoleh dari orang tua pada saat konsepsi atau selama kehamilan. Kematian neonatal dapat disebabkan oleh dua faktor, yaitu faktor ibu antara lain pelayanan kesehatan ibu hamil, infeksi ibu hamil, gizi ibu hamil, dan karakteristik dari ibu hamil (umur, paritas, dan jarak kehamilan) serta faktor janin antara lain bayi Berat Badan Lahir Rendah (BBLR), asfiksia, dan penumonia. Berikut beberapa faktor yang dapat mempengaruhi tingkat kematian bayi (Dinkes, 2018):

a. Berat Badan Lahir Rendah (BBLR)

Menurut *World Health Organization* (WHO) Bayi Berat Lahir Rendah (BBLR) didefinisikan sebagai bayi yang lahir dengan berat < 2500 gram. Berat lahir adalah berat bayi yang ditimbang dalam waktu satu jam pertama setelah lahir. Pengukuran dilakukan di tempat fasilitas (Rumah sakit, Puskesmas, dan Polindes), sedangkan bayi yang lahir di rumah waktu pengukuran berat badan dapat dilakukan dalam waktu 24 jam. BBLR dapat terjadi pada bayi kurang bulan/prematur atau disebut BBLR Sesuai Masa Kehamilan (SMK)/*Appropriate for Gestational Age* (AGA), bayi cukup bulan yang mengalami hambatan pertumbuhan selama kehamilan disebut BBLR Kecil Masa Kehamilan (KMK)/*Small for Gestational Age* (SGA) dan besar masa kehamilan/*Large for Gestational Age* (LGA).

b. Pelayanan Kesehatan Bayi

Pelayanan kesehatan bayi bertujuan untuk meningkatkan akses bayi terhadap pelayanan kesehatan dasar, mengetahui sedini mungkin jika terdapat kelainan pada bayi sehingga cepat mendapat pertolongan, pemeliharaan kesehatan dan pencegahan penyakit. Pelayanan ini ditujukan pada bayi usia 29 hari sampai dengan 11 bulan dengan memberikan pelayanan kesehatan sesuai dengan standar oleh tenaga kesehatan yang memiliki kompetensi klinis kesehatan (dokter, bidan, dan perawat) minimal 4 kali, yaitu satu kali pada umur 29 hari – 2 bulan, 1 kali pada umur 3 – 5 bulan, 1 kali pada umur 6 – 8 bulan dan 1 kali pada umur 9 – 12 bulan sesuai standar di satu wilayah kerja pada kurun waktu tertentu.

c. Bayi diberi ASI eksklusif

ASI eksklusif menurut *World Health Organization* adalah memberikan hanya ASI saja tanpa memberikan makanan dan minuman lain kepada bayi sejak lahir sampai berumur 6 bulan, kecuali obat dan vitamin.

d. Kunjungan Neonatus

Kunjungan neonatus merupakan salah satu intervensi untuk menurunkan kematian bayi baru lahir dengan melakukan Kunjungan Neonatal (KN) selama 3 (tiga) kali kunjungan yaitu Kunjungan Neonatal I (KN1) pada 6 jam sampai dengan 48 jam setelah lahir, Kunjungan Neonatal II (KN2) pada hari ke-3 sampai dengan 7 hari, dan Kunjungan Neonatal III (KN3) pada hari ke-8 sampai dengan 28 hari.

BAB III
METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari buku Profil Dinas Kesehatan Kota Makassar tahun 2017 yang dipublikasikan oleh Dinas Kesehatan Kota Makassar. Jumlah pengamatannya terdiri dari 46 puskesmas yang tersebar di 14 kecamatan di Kota Makassar. Data sekunder yang dimaksud berupa data jumlah kematian bayi, persentase bayi dengan Berat Badan Lahir Rendah (BBLR), persentase pelayanan kesehatan bayi, persentase bayi yang diberi ASI eksklusif, dan persentase kunjungan neonatus setiap puskemas di Kota Makassar seperti pada Lampiran 1.

3.2 Identifikasi Variabel

Variabel yang digunakan dalam penelitian ini terdiri dari variabel respon dan variabel prediktor. adapun variabel tersebut adalah sebagai berikut.

a. Variabel Respon

Penelitian ini menggunakan data jumlah kematian bayi yang tercatat di 46 puskesmas di Kota Makassar tahun 2017.

b. Variabel Prediktor

Variabel prediktor dalam penelitian ini terdri dari empat variabel yang diduga akan mempengaruhi jumlah kematian bayi. Keempat variabel tersebut adalah:

X_1 : Persentase Berat Badan Lahir Rendah (BBLR).

X_2 : Persentase pelayanan kesehatan bayi.

X_3 : Persentase bayi yang diberi ASI eksklusif.

X_4 : Persentase kunjungan neonatus.